



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2017

---

## **Adaptive prior weighting in generalized regression**

Held, Leonhard ; Sauter, Rafael

**Abstract:** The prior distribution is a key ingredient in Bayesian inference. Prior information on regression coefficients may come from different sources and may or may not be in conflict with the observed data. Various methods have been proposed to quantify a potential prior-data conflict, such as Box's p-value. However, there are no clear recommendations how to react to possible prior-data conflict in generalized regression models. To address this deficiency, we propose to adaptively weight a prespecified multivariate normal prior distribution on the regression coefficients. To this end, we relate empirical Bayes estimates of prior weight to Box's p-value and propose alternative fully Bayesian approaches. Prior weighting can be done for the joint prior distribution of the regression coefficients or under prior independence—separately for prespecified blocks of regression coefficients. We outline how the proposed methodology can be implemented using integrated nested Laplace approximations (INLA) and illustrate the applicability with a Bayesian logistic regression model for data from a cross-sectional study. We also provide a simulation study that shows excellent performance of our approach in the case of prior misspecification in terms of root mean squared error and coverage. Supplementary Materials give details on software implementation and code and another application to binary longitudinal data from a randomized clinical trial using a Bayesian generalized linear mixed model.

DOI: <https://doi.org/10.1111/biom.12541>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-128137>

Journal Article

Accepted Version

Originally published at:

Held, Leonhard; Sauter, Rafael (2017). Adaptive prior weighting in generalized regression. *Biometrics*, 73(1):242-251.

DOI: <https://doi.org/10.1111/biom.12541>

## Adaptive prior weighting in generalized regression

Leonhard Held and Rafael Sauter

Department of Biostatistics

Epidemiology, Biostatistics and Prevention Institute

University of Zurich

Hirschengraben 84, 8001 Zurich, Switzerland

Email: [leonhard.held@uzh.ch](mailto:leonhard.held@uzh.ch), [rafael.sauter@uzh.ch](mailto:rafael.sauter@uzh.ch)

**SUMMARY:** The prior distribution is a key ingredient in Bayesian inference. Prior information on regression coefficients may come from different sources and may or may not be in conflict with the observed data. Various methods have been proposed to quantify a potential prior-data conflict, such as Box's  $p$ -value. However, there are no clear recommendations how to react to possible prior-data conflict in generalized regression models. To address this deficiency, we propose to adaptively weight a pre-specified multivariate normal prior distribution on the regression coefficients. To this end, we relate empirical Bayes estimates of prior weight to Box's  $p$ -value and propose alternative fully Bayesian approaches. Prior weighting can be done for the joint prior distribution of the regression coefficients or - under prior independence - separately for pre-specified blocks of regression coefficients. We outline how the proposed methodology can be implemented using integrated nested Laplace approximations (INLA) and illustrate the applicability with a Bayesian logistic regression model for data from a cross-sectional study. We also provide a simulation study that shows excellent performance of our approach in the case of prior misspecification in terms of root mean squared error and coverage. Supplementary Materials give details on software implementation and code and another application to binary longitudinal data from a randomized clinical trial using a Bayesian generalized linear mixed model.

**KEY WORDS:** Hyper- $g$  prior; Prior weight; prior-data conflict;  $g$ -prior; generalized regression; INLA

This paper has been submitted for consideration for publication in *Biometrics*

## 1. Introduction

Appropriate specification of the prior distribution is a key ingredient in Bayesian statistics. It is also considered as the most controversial feature of Bayesian inference. In this paper we discuss the role of the prior distribution in generalized regression models from a novel perspective. We consider a commonly used setup where a proper multivariate normal prior is assigned to the regression coefficients. Prior weighting is achieved by a scalar  $g > 0$ , acting multiplicatively on the prior covariance matrix. Thus, the prior weight is represented by the inverse  $w = 1/g$ . The focus of this paper will be on empirical and fully Bayesian approaches to estimate the inverse prior weight  $g$  from the data at hand. This is different from the Evans and Jang (2011b) approach to identify a whole set of weakly informative priors relative to a pre-specified base prior, since their procedure is not data dependent, *i. e.* not adaptive.

We distinguish four different sources for a prior distribution. First, prior information may come from historical data of the same structure as the current data. For example, data from past clinical trials may be used to construct a suitable prior for the analysis of data from a current trial with the same outcome. Approaches to integrate historical data include the robust meta-analytic approach (Schmidli et al., 2014) and the power prior (Ibrahim and Chen, 2000; Duan et al., 2006; Neuenschwander et al., 2009), which introduces a weight parameter to discount historical data.

Secondly, the prior distribution may come from elicitation of expert opinion (O’Hagan et al., 2006; Spiegelhalter et al., 2004, Section 5.2). For example, Miettinen et al. (2008) develop a risk prediction model for the presence of pneumonia, elicited from 22 clinical experts. This model has been subsequently updated in Held et al. (2012) using data on more than 600 patients presenting with cough and fever at a general practitioner’s practice in Switzerland.

However, historical data or expert opinion may not be available for the problem at hand, but an informative prior may still be warranted based on contextual reasoning. Greenland (2006, 2007) argues strongly that proper priors should be used in the analysis of epidemiological studies to avoid implicit unrealistic assumptions of the corresponding frequentist analysis (operationally equivalent to a Bayesian analysis with improper priors on the parameters of interest). For example, Greenland (2006) specifies a normal prior with mean zero and variance 1/2 for a log odds ratio parameter to reflect the prior belief that the median odds ratio is 1 and the odds ratio is between 1/4 and 4 with 95% probability *a priori*. Other choices for prior mean and variance are possible, of course, and Greenland (2006) recommends to perform a sensitivity analysis by varying the prior variance.

Fourthly and finally, proper default prior distributions may be used as a conservative guess or to avoid the problem of diverging maximum likelihood estimates in logistic regression due to complete separation (Albert and Anderson, 1984). For example, the ridge prior, a multivariate normal prior with mean zero and covariance matrix proportional to the identity matrix, is a commonly used default prior. Zellner's *g*-prior for linear models (Zellner, 1986), with prior covariance matrix proportional to the covariance matrix of the maximum likelihood estimate (MLE) of the regression coefficients, is another default prior, which has the attractive feature that *g* can be interpreted as relative inverse prior sample size, see for example Marin and Robert (2007, Section 3.2.2) or Liang et al. (2008). The *g*-prior is a natural approach to incorporate prior correlations between regression coefficients and automatically adjusts for different variances of the covariates. Suitable extensions of the *g*-prior to generalized linear models (GLMs) are discussed in Sabanés Bové and Held (2011). Both ridge and *g*-priors are often used for Bayesian model selection, where the prior distribution needs to be proper to ensure that

the marginal likelihood is well-defined and the corresponding Bayes factors (Kass and Raftery, 1995) can be calculated.

Methodology to estimate the inverse prior weight  $g$  goes back to the literature on ridge regression (Lindley and Smith, 1972; Hoerl et al., 1975; Box, 1980). Empirical Bayes (EB) estimates for  $g$ -priors have been proposed by Copas (1983) both for the linear and the logistic regression model. Fully Bayesian (FB) approaches to estimate  $g$  have been advocated in the linear model with regression splines (Denison et al., 2002, Section 3.5 and references in Section 3.8), using an inverse gamma hyperprior for  $g$  in combination with a ridge prior. The support of the inverse gamma distribution is the whole positive real line, thus the prior weight can be either de- or increased. Prior distributions for the parameter  $g$  of the  $g$ -prior have been proposed in Cui and George (2008); Liang et al. (2008) and Held et al. (2015) in the context of Bayesian model selection.

This paper is structured as follows. In the generalized linear model with a multivariate normal prior on the regression coefficients (Section 2) we first discuss methodology originally proposed by Box (1980) to quantify the prior-data conflict, see also Spiegelhalter et al. (2004, Section 5.8), Greenland (2006) and Evans and Moshonov (2006). We then proceed and describe methods to estimate the prior weight, represented by the parameter  $1/g$ . This leads to *adaptive* prior weighting, as opposed to approaches with fixed prior weight. We review empirical Bayes procedures (Copas, 1983, 1997) to estimate  $g$  in the  $g$ -prior setting and extend those to any normal prior. Furthermore, we show that EB estimates of  $g$  correspond to intermediate solutions between prior-data agreement and disagreement. We finally propose fully Bayes procedures to estimate  $g$  using a suitable hyperprior for  $g$ . If blocks of regression coefficients are *a priori* independent, then the approach can be extended to separately weight each block. Application in more complex regression models is also possible, for example in generalized linear mixed

models. Inference for Bayesian GLMs with a hyperprior on  $g$  is done using integrated nested Laplace approximations (INLA) (Rue et al., 2009), to avoid the commonly used computer-intensive Markov chain Monte Carlo (MCMC) simulation.

Two applications are considered in this paper: A Bayesian logistic regression model for data from a cross-sectional study (Sullivan and Greenland, 2013) is described in Section 3.1, while a Bayesian analysis of binary longitudinal data with a generalized linear mixed model is outlined in Supplementary Materials. Section 3.2 describes additional simulation studies that have been performed to investigate the properties of the proposed methodology. We close with some discussion in Section 4.

## 2. Methodology

Consider a generalized linear model (GLM) with outcomes  $y_i$ ,  $i = 1, \dots, n$ , and linear predictor  $\eta_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta}$ , where the vector of regression coefficients  $\boldsymbol{\beta}$  has dimension  $d$ . The mean  $\mu_i = h(\eta_i)$  of  $y_i$  is obtained with the response function  $h(\eta_i)$ , the variance function  $v(\mu_i)$  determines the variance of  $y_i$ . We use a Gaussian prior with mean  $\boldsymbol{\nu}$  and covariance matrix  $g\boldsymbol{\Sigma}$  for  $\boldsymbol{\beta}$ , *i.e.*  $\boldsymbol{\beta} \sim N(\boldsymbol{\nu}, g\boldsymbol{\Sigma})$ . The intercept  $\alpha$  can be extremely sensitive to how covariates are centered and how factors are coded, so we follow the recommendations by Greenland and Mansournia (2015, Section 7) and use Jeffreys' prior  $f(\alpha) \propto 1$ . More informative priors may induce unjustifiable shrinkage of the intercept towards an arbitrary prior mean. We note that also Gelman et al. (2008) use an extremely dispersed Cauchy prior for the intercept, negligibly different from our flat prior.

### 2.1 Prior-data conflict

Box (1980) has suggested an approach to quantify a potential conflict between the prior distribution and the observed data. The methodology is based on the prior predictive density  $f(\mathbf{y})$  of the data  $\mathbf{Y}$  and compares the distribution of  $f(\mathbf{Y})$  with  $f(\mathbf{y} = \mathbf{y}_{\text{obs}})$ , here

$\mathbf{y}_{\text{obs}}$  denotes the observed data. Specifically, Box's  $p$ -value is defined as the probability

$$\Pr\{f(\mathbf{Y}) \leq f(\mathbf{y}_{\text{obs}})\}, \quad (1)$$

where a small value of (1) implies that the observation  $\mathbf{y}_{\text{obs}}$  has relatively low prior predictive density, *i.e.* indicates prior-data conflict. To avoid some anomalous behavior, Evans and Moshonov (2006) proposed to replace the data  $\mathbf{Y}$  in (1) with a minimal sufficient statistic for the parameter of interest. This ensures that the method provides a measure of prior-data conflict only, and not a confounded check of the model + prior combination. In more recent work, Evans and Jang (2011a) show the consistency of the Evans and Moshonov (2006) methodology and discuss the lack of invariance of the original Box (1980) approach.

However, exact computation of (1) is difficult in GLMs. Therefore, Greenland (2006) suggested to consider the MLE  $\hat{\beta}_{\text{ML}}$  (of course a minimal sufficient statistic for  $\beta$ ) as the “data” with (asymptotic)  $\hat{\beta}_{\text{ML}} | \beta \sim N(\beta, \mathcal{T})$  distribution (Fahrmeir and Kaufmann, 1985), where  $\mathcal{T}$  denotes the (estimated) covariance matrix of the MLE. Combining this with a  $N(\nu, g\Sigma)$  prior for  $\beta$  gives the (approximate) prior predictive distribution  $\hat{\beta}_{\text{ML}} \sim N(\nu, \mathcal{T} + g\Sigma)$ . The standardized difference

$$T(g) = (\hat{\beta}_{\text{ML}} - \nu)^\top (\mathcal{T} + g\Sigma)^{-1} (\hat{\beta}_{\text{ML}} - \nu) \quad (2)$$

can then be evaluated against a  $\chi^2$ -distribution with  $d$  degrees of freedom to compute Box's  $p$ -value. This approximates the predictive check by Box (1980, eq. (3.9)) for the linear model based on the  $F$ -distribution using an additional improper prior  $f(\sigma^2) \propto \sigma^{-2}$  on the residual variance.

## 2.2 Estimates of prior weight

In the absence of prior information on  $\Sigma$ , the generalized  $g$ -prior (Sabanés Bové and Held, 2011) can be used as default. The corresponding prior covariance matrix is taken as  $\Sigma = c(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1}$  where  $\mathbf{W}$  is a diagonal matrix with corresponding weights on the

diagonal (*e.g.* the binomial sample sizes for logistic regression). Here, the columns of the design matrix  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$  are assumed to be centred, *i.e.*  $\mathbf{X}^\top \mathbf{W} \mathbf{1} = \mathbf{0}$ . The constant  $c = c(\alpha)$  depends on the specific GLM and is defined as

$$c(\alpha) = v(h(\alpha))\{h'(\alpha)\}^{-2} \quad (3)$$

where  $h'(\cdot)$  is the derivative of  $h(\cdot)$ , see Copas (1983) for a derivation for binary outcomes and Sabanés Bové and Held (2011) for a general treatment. The implied shrinkage of the MLE  $\hat{\boldsymbol{\beta}}_{\text{ML}}$  is then approximately as for the standard  $g$ -prior in the linear model (Liang et al., 2008) with posterior mean

$$\mathbb{E}(\boldsymbol{\beta} | \mathbf{y}) \approx \left( \frac{n \cdot \hat{\boldsymbol{\beta}}_{\text{ML}} + n/g \cdot \boldsymbol{\nu}}{n + n/g} \right),$$

which reduces to  $\mathbb{E}(\boldsymbol{\beta} | \mathbf{y}) \approx \frac{g}{g+1} \hat{\boldsymbol{\beta}}_{\text{ML}}$  for  $\boldsymbol{\nu} = \mathbf{0}$  (Held et al., 2015). Thus  $t = g/(g+1)$  can be interpreted as shrinkage factor for the generalized  $g$ -prior with prior mean  $\mathbf{0}$ .

To derive an EB estimate of  $g$ , we note that we can re-write (2) with  $\mathcal{T} \approx c(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1}$  (Copas, 1983) as

$$T(g) = \frac{1}{1+g} \frac{1}{c} (\hat{\boldsymbol{\beta}}_{\text{ML}} - \boldsymbol{\nu})^\top (\mathbf{X}^\top \mathbf{W} \mathbf{X}) (\hat{\boldsymbol{\beta}}_{\text{ML}} - \boldsymbol{\nu}). \quad (4)$$

Equating (4) with its expectation  $d$  (subject to  $g \geq 0$ ) gives the analytic solution

$$\begin{aligned} \hat{g} &= \max \left\{ \frac{1}{d} \frac{1}{c} (\hat{\boldsymbol{\beta}}_{\text{ML}} - \boldsymbol{\nu})^\top (\mathbf{X}^\top \mathbf{W} \mathbf{X}) (\hat{\boldsymbol{\beta}}_{\text{ML}} - \boldsymbol{\nu}) - 1, 0 \right\} \\ &\approx \max \{ z_{\text{obs}}/d - 1, 0 \}, \end{aligned} \quad (5)$$

here  $z_{\text{obs}}$  denotes the observed deviance (relative to the null model with  $\boldsymbol{\beta} = \mathbf{0}$ ), obtained from fitting a standard GLM to the data at hand (Copas, 1983, 1997). By construction, plugging-in (5) into (4) gives (for  $\hat{g} > 0$ )  $T(\hat{g}) = d$ , so Box's  $P$ -value can be easily evaluated for the adapted  $N(\boldsymbol{\nu}, \hat{g} \boldsymbol{\Sigma})$  prior with EB estimate  $\hat{g}$ . Box's  $P$ -value turns out to be 0.32, 0.37, 0.39  $\rightarrow$  0.5 for increasing degrees of freedom  $d = 1, 2, 3 \rightarrow \infty$ . This illustrates that in regular cases (where  $\hat{g} > 0$ ) empirical Bayes automatically adjusts the



weight parameter  $g$  based on the compatibility between the prior and the data and leads to unremarkable Box's  $P$ -values between 0.32 and 0.5.

The approach can be easily extended to arbitrary prior mean  $\nu$  if we evaluate the deviance not against the null model  $\nu = \mathbf{0}$  but against a model with non-zero prior mean  $\nu$ . This can be achieved by fitting a GLM with offset  $X\nu$ . For arbitrary prior covariance matrix  $\Sigma$  an empirical Bayes-type (moment-based) estimate of  $g$  can be implemented by equating (2) with the mean  $d$  of the  $\chi^2(d)$ -distribution and numerically solving for  $g$ .

The empirical Bayes approach avoids arbitrary choices of  $g$  which may be at odds with the data. However, the uncertainty about the estimate  $\hat{g}$  is ignored, *i. e.* the estimate  $\hat{g}$  is treated as the true value  $g$ . This is particularly worrying if  $\hat{g} = 0$ , since then the posterior of  $\beta$  degenerates to a point mass at the prior mean  $\nu$ , no matter what the data are. In contrast, a fully Bayesian approach to estimate  $g$  will incorporate the uncertainty about the estimate from its posterior distribution. If the prior distribution comes from historical data, a beta prior is commonly used for  $1/g$ , which restricts the range of  $g$  to values larger than unity (Duan et al., 2006). The prior can therefore only be down- but not up-weighted. However, if the prior distribution is not based on historical data, then also increasing the weight of the prior distribution may be warranted by the data at hand. We will illustrate this in Application 3.1.

For Bayesian model selection based on the  $g$ -prior, Liang et al. (2008) suggest to use the hyper- $g$  prior with prior density

$$f(g) = \frac{a-2}{2}(1+g)^{-a/2} \quad (6)$$

for  $g$ , which is proper for  $a > 2$ . This prior distribution is a special case of a class of prior distributions proposed by Cui and George (2008) and induces a beta distribution for the shrinkage factor  $t = g/(g+1)$ :  $t \sim \text{Be}(1, a/2 - 1)$ . Of particular interest is the case  $a = 4$ , where the prior on the shrinkage factor  $t$  is standard uniform and thus the prior median

of  $g$  is 1. Furthermore, the distribution of  $w = 1/g$  is the same as the distribution of  $g$ , *i.e.* the prior has no preference regarding up- or down-weighting. The cdf of  $g$  has a simple analytic form,  $F(g) = g/(g+1)$ , so prior probabilities of interest can be easily calculated, *e.g.*  $\Pr(1/2 \leq g \leq 2) = 1/3$  or  $\Pr(1/19 \leq g \leq 19) = 0.9$ . This “standard” hyper- $g$  prior has infinite expectation, so is sufficiently dispersed. Furthermore, under the generalized  $g$ -prior a uniform prior on the shrinkage factor  $t$  implies that the posterior mode of the shrinkage factor  $t = g/(g+1)$  is asymptotically equal to the corresponding EB estimate based on (5) (Held et al., 2015). Thus, the standard hyper- $g$  prior regularizes empirical Bayes and can be considered as a natural choice for a hyperprior for  $g$ .

An alternative symmetric prior would be  $f(g) = \pi^{-1}g^{-0.5}(1+g)^{-1}$ , which corresponds to  $t \sim \text{Be}(1/2, 1/2)$ , the so-called horseshoe prior (Carvalho et al., 2010). This choice is also indifferent regarding up- or down-weighting, but puts substantially more prior mass to extreme values of  $g$ . For example, under the horseshoe prior  $\Pr(1/161 \leq g \leq 161) \approx 0.9$ . The Strawderman-Berger (short Strawderman) prior (Berger, 1980), obtained from (6) with  $a = 3$ , places more weight on larger values of  $g$ , *i.e.* treats  $g$  not symmetric. For example, the prior median is now 3. Another non-symmetric prior on  $g$  is the inverse gamma distribution. This choice is often made for convenience due to conjugacy in the normal linear model (Denison et al., 2002, Section 3.3), but lacks a deeper motivation as a suitable prior for a weight parameter.

Whatever prior for  $g$  is used, calculation of the posterior distribution can be done using numerical integration with INLA (Rue et al., 2009), which we describe in Section 2.3.

### 2.3 Implementation in INLA

The traditional choice to implement the proposed approach would be Markov chain Monte Carlo (MCMC), but we prefer a numerical approach based on INLA to avoid potential convergence problems and the associated Monte Carlo error of MCMC. How-

ever, adaptive prior weighting with the R-INLA interface (see Martins et al. (2013, Section 2.3) for a summary) requires specific amendments to the model which we now briefly describe. More details can be found in Supplementary Materials.

Consider a GLM as described in Section 2 with linear predictor  $\eta_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta}$  where  $\boldsymbol{\beta} \sim N(\boldsymbol{\nu}, g \boldsymbol{\Sigma})$  *a priori*. R-INLA does not allow to specify this prior directly, so the formulation needs to be re-written as  $\eta_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\nu} + \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}}$  where  $\tilde{\boldsymbol{\beta}} \sim N(\mathbf{0}, g \boldsymbol{\Sigma})$ . Therefore,  $o_i = \mathbf{x}_i^\top \boldsymbol{\nu}$  can be used as offset variable and it is sufficient to consider priors for the regression coefficients  $\tilde{\boldsymbol{\beta}}$  with mean zero. The idea is to treat the mean-zero regression coefficients  $\tilde{\boldsymbol{\beta}}$  as a Gaussian Markov random field (GMRF) (Rue and Held, 2005) with pre-specified precision matrix  $\boldsymbol{\Sigma}^{-1}$  - up to the possibly unknown multiplicative weight factor  $w = 1/g$ . However, it is not possible to directly compute the product  $\mathbf{x}_i^\top \tilde{\boldsymbol{\beta}}$  for all observations  $i = 1, \dots, n$ . The trick is now to use the copy feature (Martins et al., 2013, Section 4.3) in order to define  $d$  identical copies of  $\tilde{\boldsymbol{\beta}}$  in the model formulation. The  $j$ -th component of the  $j$ -th copy of  $\tilde{\boldsymbol{\beta}}$  is eventually multiplied with the corresponding covariate values  $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^\top$ ,  $j = 1, \dots, d$ .

The R-INLA default treats the weight  $w = 1/g$  as unknown and uses a gamma hyperprior for it. However, the software allows the user to define any suitable prior density, either as an expression or in tabulated form, giving the value of the prior density on a suitable grid. Note that INLA requires this to be done for  $\log(w)$ . Here we have used the tabulated approach for the standard hyper-g, horseshoe and Strawderman prior and have computed the corresponding density of  $\log(1/g)$  with a change-of-variables. The weight  $w$  can also be fixed at a pre-specified value which we have used for the case  $g = 1$ .

The described implementation can be generalized to independent blocks of regression coefficients  $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p$ , say, in order to weight the blocks with different weight parameters

$g_1, \dots, g_p$ , see end of Section 3.1 for an example. It is also straightforward to estimate the weight of the prior on the regression coefficients  $\beta$  in more complex models such as generalized linear mixed models, see Supplementary Materials.

### 3. Applications

#### 3.1 Bayesian analysis of a logistic regression model

Sullivan and Greenland (2013) consider data from a cross-sectional study on obstetric care and neonatal death at a teaching hospital, originally analysed in Neutra et al. (1978). The binary outcome variable (death yes/no) is related to 14 explanatory variables. They are all binary with frequencies between 0.3% (variable hydram) and 77% (variable nomonit). A logistic model with all 14 variables converges without problem, but produces some extremely inflated estimates due to the sparseness of the data with only 17 deaths observed among  $n = 2992$  births. In order to “shrink the estimates back to something more reasonable”, Sullivan and Greenland (2013) select an informative prior for the regression coefficients  $\beta$ . More specifically,  $\beta$  is assumed to be *a priori* normally distributed with mean  $\nu_{SG} = \log(\mathbf{OR})$ , where the vector  $\mathbf{OR} = (2, 2, 2, 4, 2, 1, 4, 2, 2, 2, 4, 2, 2, 4)^\top$  contains the prior median odds ratios for each explanatory variable. The prior covariance matrix  $\Sigma$  has been chosen to be diagonal with all variances equal to  $1/2$ . The intercept is considered as a nuisance parameter with improper (flat) prior. The implementation of this formulation using INLA is described in Supplementary Materials.

There is no evidence for prior-data disagreement for the informative  $N(\nu_{SG}, \Sigma)$  prior (equation (2) gives Box’s  $p=0.91$ ). Even if the prior mean  $\nu$  would be set to zero, *i. e.* for a  $\beta \sim N(\mathbf{0}, \Sigma)$  prior, there would be no compelling evidence for prior-data conflict (Box’s  $p=0.13$ ). The EB estimates of  $g$  are  $\hat{g} = 0.00$  and  $\hat{g} = 2.10$  in these two cases. Box’s  $p$ -values using the EB estimates of  $g$  in the adapted prior covariance matrix  $\hat{g} \Sigma$  are  $p=0.60$

and  $p=0.45$ , respectively, so in both cases close to 0.5, as expected from the discussion in Section 2.2.

If we combine the Sullivan and Greenland (2013) prior with the standard hyper- $g$  prior, the resulting posterior for  $g$  (see Figure 1) has median 0.16 (equi-tailed 95% credible interval (CI): 0.01 to 0.81). Thus, the hyper- $g$  prior increases the weight of the prior on the regression coefficients by a median factor of  $1/0.16 \approx 6.3$ , which corresponds to a reduction of the prior variance from 0.5 to 0.08. There is quite large posterior uncertainty regarding  $g$ , but values larger than 1 are rather unlikely, see Figure 1.

[Figure 1 about here.]

Sullivan and Greenland (2013) pay particular attention to one explanatory variable, hydram ( $x_7$ ), an indicator of hydramnios during pregnancy, with MLE 60 (95% Wald confidence interval 5.7 to 635, 95% profile likelihood confidence interval 2.8 to 478) of the corresponding odds ratio  $\exp(\beta_7)$ . The Sullivan and Greenland (2013) prior for the corresponding log odds ratio  $\beta_7$  is normal with mean  $\log(4)$  and variance  $1/2$ , resulting in a prior median odds ratio of 4 with equi-tailed 95% prior CI from 1 to 16. Combining this prior with the data gives posterior median 6.1 (95% CI: 1.6 to 22.8) for the odds ratio.

If we treat  $g$  as unknown with hyper- $g$  prior, then the posterior median of  $\exp(\beta_7)$  is 4.3 (95% CI: 2.3 to 10.5). Although the hyper- $g$  prior implies a substantially more dispersed marginal prior on the odds ratio (95% prior CI: 0.21 to 75.0), the posterior is actually narrower than for fixed  $g = 1$ . The corresponding OR estimates under the hyper- $g$ , horseshoe and Strawderman prior are given in Table 1, together with DIC values to assess the model fit (Spiegelhalter et al., 2002). The posterior distributions of the corresponding regression coefficient  $\beta_7$  (*i.e.* the log odds ratio) are compared in Figure 2. Of note, the posterior under the horseshoe prior is substantially more peaked and narrower than under the hyper- $g$  and Strawderman prior. The model fit turns out to be 3-4 units better

for the three fully Bayesian approaches compared to the analysis with fixed  $g = 1$ , with the horseshoe prior having the lowest DIC value.

[Figure 2 about here.]

The same analysis has been conducted with prior mean  $\nu = \mathbf{0}$ , keeping the diagonal prior covariance matrix  $\Sigma$  with entries  $1/2$ . The resulting posterior of  $g$  has median 1.20 (95%-CI: 0.46 to 3.13) and is also displayed in Figure 1. Thus, the hyper- $g$  prior now decreases the weight of the prior on the regression coefficients, but only slightly by a median factor of  $1/1.20 = 0.83$ , with considerable uncertainty regarding  $g$ . Accordingly, the posterior distribution of the regression coefficient related to the variable *hydram* now barely differs whether we use fixed  $g = 1$  or the hyper- $g$  prior, see Figure 2. Indeed, the posterior median of  $\exp(\beta_7)$  is now 1.6 (95% CI: 0.4 to 6.3) for  $g = 1$  and 1.8 (95% CI: 0.4 to 13.4) for the hyper- $g$  prior. Estimates for the other two priors (horseshoe and Strawderman-Berger) are given in Table 1, again with DIC values, which are now very similar for the different approaches.

Figure 1 also shows prior and posterior of  $t = g/(g + 1)$ . For the prior mean  $\nu_{SG}$ , the posterior mode of  $t$  is close to zero, as expected from the EB estimate  $\hat{g} = 0$ . For prior mean  $\mathbf{0}$ , the posterior mode of  $t$  is 0.55, reasonably close to the corresponding EB estimate  $\hat{g} = 2.10/(1 + 2.10) = 0.68$ . The small difference can be explained by substantial non-normality of the posterior distribution of  $\beta$  as already indicated by the difference between the Wald and the profile likelihood confidence interval for  $\beta_7$  given above. Thus, the assumed posterior normality of the MLE to compute the EB estimate  $\hat{g}$  based on (2) does not hold, whereas the fully Bayesian approach does not require this assumption.

One could argue that the above change to the prior mean  $\nu = \mathbf{0}$  should be accompanied by a more flexible formulation for the prior variances. To do so, we now introduce a new prior “block hyper- $g$ ” formulation with three different  $g$  parameters:  $g_1$  for the block of

nine covariates with original prior mean of  $\log(2)$ ,  $g_2$  for the block of four covariates with prior mean of  $\log(4)$ , and  $g_3$  parameter for the single covariate with prior mean 0. Thus the prior weight is now allowed to vary from block to block.

The posterior median of  $g_1$  is 0.19 (95% CI: 0.02 to 1.4),  $g_2$  has posterior median 4.1 (95% CI: 0.9 to 20.9), while  $g_3$ , the inverse prior weight of the single covariate with prior mean 0, has posterior median 0.64 with large posterior uncertainty (95% CI: 0.04 to 9.6). Thus, the weight of the prior distribution has been increased by a median factor of  $1/0.19 \approx 5.3$  for the first block of parameters, whereas the weight of the second block (which includes the variable `hydram`) has been decreased by a median factor of 4.1. The posterior median of  $\exp(\beta_7)$  is now 5.5 (95% CI: 0.4 to 95.5). Thus, the decreased weight of the prior distribution leads to a substantially larger OR estimate and a decreased precision of the regression coefficient, compared to the analysis with one unknown weight parameter  $1/g$ . Of note, this formulation gives the best model fit with DIC value 182.6, see Table 1.

[Table 1 about here.]

### 3.2 Simulation studies

In a simulation study we have compared our approach with different hyperpriors for  $g$  (including fixed  $g = 1$ ) and different degrees of misspecification of the prior mean (Section 3.3) or the covariance matrix (Section 3.4). To do so, we simulate  $\beta$  from a (possibly misspecified) “prior” distribution and subsequently  $y$  from a logistic regression model with linear predictor  $\alpha + X^\top \beta$ , here  $X$  is the same design matrix as in the application described in Section 3.1. For the subsequent analyses with INLA we use a normal prior for  $\beta$  with mean  $\nu_{SG}$  and covariance matrix  $g \Sigma$  where  $\Sigma = \text{diag}(0.5, \dots, 0.5)$ .

### 3.3 Simulation study I with shifted mean

Misspecification of the prior mean  $\nu_{\text{SG}}$  is achieved by adding a shift parameter  $\epsilon_s$ ,  $s = 1, \dots, S$  to each component of  $\nu_{\text{SG}}$ . Here we use  $\epsilon_s \in \epsilon = (-2.6, -2.4, \dots, 0, \dots, 2.4, 2.6)$ , so  $S = 27$ , and sample  $\beta_s^{(k)}$ ,  $k = 1, \dots, K = 1000$  from  $N(\nu_{\text{SG}} + \epsilon_s, \Sigma)$ . For each  $\beta_s^{(k)}$  we compute the linear predictor  $\eta_s^{(k)} = \alpha + \mathbf{X}^\top \beta_s^{(k)}$  and the risk probability vector  $\pi_s^{(k)} = \frac{\exp(\eta_s^{(k)})}{1 + \exp(\eta_s^{(k)})}$  and finally generate binary response vectors  $\mathbf{y}_s^{(k)} \sim \text{Bin}(\pi_s^{(k)}, 1)$ . To avoid problems with complete separation, the intercept  $\alpha$  has been chosen such that the proportion of events ( $y = 1$ ) is close to 0.5. The simulated data  $(\mathbf{y}_s^{(k)}, \mathbf{X})$  are now analysed with R-INLA using a  $N(\nu_{\text{SG}}, g \Sigma)$  prior for  $\beta$  and the following priors on  $g$ : Fixed  $g = 1$ , Hyper- $g$ , horseshoe and Strawderman-Berger.

In total,  $27 \times 1000 \times 4 = 108\,000$  calls of R-INLA have been made. Except for fixed  $g = 1$ , the posterior median of  $g$  has been computed and averaged across  $K = 1000$  analyses for each  $\epsilon_s$ . This is shown in the top row of Figure 3. One can see how the different approaches react to prior misspecification with increasing estimates of  $g$  for increasing  $|\epsilon_s|$ . There are only minor differences between the different approaches with a slight bias towards  $g > 1$  of the Strawderman prior in the case of no misspecification ( $\epsilon_s = 0$ ).

The next row in Figure 3 gives the root mean squared error (RMSE)

$$\sqrt{\frac{1}{K} \sum_{k=1}^{1000} \left\{ E(\beta_j | \mathbf{y}_s^{(k)}) - \beta_{s,j}^{(k)} \right\}^2}$$

between the posterior mean of  $\beta_j$  and the true value, the  $j$ -th component  $\beta_{s,j}^{(k)}$  of  $\beta_s^{(k)}$ . Shown are the results for two covariates, nullip with a balanced proportion of 49% “cases” and hydram with only 0.3% cases. The third row gives the corresponding mean posterior standard deviation (MPSD) of the two covariates. We show only results for the hyper- $g$  approach, since the other two priors gave virtually identical results. It is interesting to see how the hyper- $g$  approach reacts to model misspecification with much



lower RMSE and larger MPSD in the case of model misspecification. As one would expect, differences between hyper-g and fixed-g (both in terms of RMSE and MPSD) increase with increasing amount of misspecification.

Finally, the last row in Figure 3 gives the coverage of equi-tailed 95% credible intervals for the components of  $\beta$ , averaged across all 14 covariates. Whereas the hyper-g and the horseshoe prior (Strawderman gives very similar results) have coverage very close to the nominal 95% level, the empirical coverage of the fixed-g analysis drops quickly to values of 85% and below.

[Figure 3 about here.]

### 3.4 Simulation study II with scaled covariance matrix

In a second simulation study, we have investigated the effect of misspecification of the prior covariance matrix. The study has been conducted as in Section 3.4, with the only difference that  $\beta$  is now generated from a  $N(\nu_{SG}, \delta_s \Sigma)$  distribution where the  $S = 31$  components of  $\delta = (1/20, \dots, 1, \dots, 20)$ , equally-spaced on the log-scale, quantify the amount of prior misspecification.

The results are shown in Figure 4. The adaptive approaches react to prior misspecification with smaller values of the posterior medians of  $g$  for small values of  $\delta_s$  and *vice versa*. The RMSE of the fixed-g approach increases dramatically for larger  $\delta_s$ , whereas the increase of the hyper-g approach is only moderate. Of note, the MPSD of the hyper-g is now smaller than for the fixed-g approach for small of  $\delta_s$ . As in Section 3.3, the coverage of the 95% credible intervals is very close to the nominal level for the hyper-g approach (Strawderman again not shown, since visually indistinguishable). The horseshoe prior gives similar results, but with coverage slightly too low for small values of  $\delta_s$ . The fixed-g approach has coverage too high for small values of  $\delta_s$  and coverage too low for larger values of  $\delta_s$ .

[Figure 4 about here.]

#### 4. Discussion

We have proposed a novel approach to update the weight of the prior distribution in the light of the current data. We have focused on the common scenario where the prior distribution for the regression coefficients is multivariate normal. Adaptive prior weighting is achieved by introducing an unknown multiplicative scalar  $g$  for the prior covariance matrix. A hyperprior for  $g$  allows to estimate the weight of the prior distribution in the light of the current data. The application showed that the hyper- $g$  prior allows for both up- or down-weighting of the weight of the prior distribution, so the adapted prior is not necessarily weakly informative relative to the original prior (Evans and Jang, 2011b).

Prior information on regression coefficients from historical data can often be assumed to be normal due to the approximate normality of the posterior distribution, *e. g.* Bernardo and Smith (2000). A normal prior distribution is therefore a natural choice. The explicit incorporation of a prior weight parameter in our approach can be used to inform researchers on the appropriateness of the original prior being used. The simulation study has shown that the posterior distribution of  $g$  informs appropriately about possible misspecification of the prior distribution.

However, if the interest is primarily in the regression coefficients, an alternative way to interpret a hyperprior on the inverse prior weight parameter  $g$  is to consider the implied marginal prior distribution on the regression coefficients, a scale mixture of normals (West, 1987). The prior weight  $w = 1/g$  is then considered a nuisance parameter and its posterior distribution is only of secondary interest. For example, an inverse gamma hyperprior for  $g$  leads to a “robust” Student  $t$  rather than a normal prior distribution for the regression coefficients. As a special case one obtains a Cauchy prior as proposed in Gelman et al. (2008) for logistic regression. From that perspective, our approach can

be viewed as replacing a normal prior on the regression coefficients with a “robustified” scale mixture of normals prior. In contrast to Student  $t$  or Cauchy priors, the hyper- $g$  prior on  $g$  allows for a symmetric treatment of the weight parameter and can be viewed as a regularized version of empirical Bayes, thus balancing prior-data agreement and disagreement. However, the implied marginal distribution of the regression coefficients does not have a standard form (Liang et al., 2008).

As an extension of our approach we have introduced several independent weight parameters for blocks of regression coefficients in the application described at the end of Section 3.1. In the limit, every regression coefficient can have its own weight parameter as long as the prior covariance matrix is diagonal. However, the data do often not contain enough information to allow precise estimation of all the  $g_j$ 's, see also Denison et al. (2002, Section 3.8).

The advantage of the implementation in INLA is that the proposed methodology can easily be applied in more complex models, *e.g.* generalized linear mixed models. In future work we also plan to compare the sensitivity of the posterior of the regression coefficients (with and without adaptive prior weighting) with respect to mean and covariance matrix of the normal prior (Roos and Held, 2011; Roos et al., 2015).

## Supplementary Materials

Implementation details and applications referenced in Sections 2 and 3 as well as data and code are available with this paper at the Biometrics website on Wiley Online Library.

## Acknowledgments

R-code by Daniel Sabanés Bové, written for the analysis described in Held et al. (2012), was helpful to implement the approach proposed in this paper. We acknowledge helpful

and constructive comments by the Editor Yi-Hau Chen, an Associate Editor and two referees on an earlier version of this article.

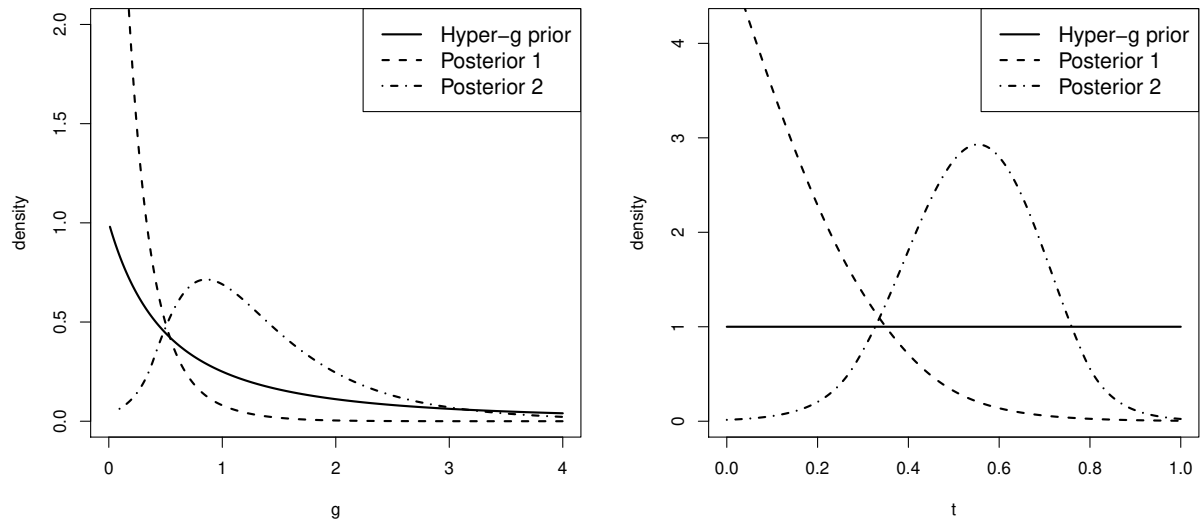
## References

- Albert, A. and Anderson, J. A. (1984). On the existence of maximum-likelihood estimates in logistic regression models. *Biometrika* **71**, 1–10.
- Berger, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Annals of Statistics* **8**, 716–761.
- Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.
- Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)* **143**, 383–430.
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–480.
- Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society. Series B (Methodological)* **45**, 311–354.
- Copas, J. B. (1997). Using regression models for prediction: shrinkage and regression to the mean. *Statistical Methods in Medical Research* **6**, 167–183.
- Cui, W. and George, E. I. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference* **138**, 888–900.
- Denison, D. G. T., Holmes, C. C., Mallick, B. K., and Smith, A. F. M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Wiley Series in Probability and Statistics. Wiley, Chichester.
- Duan, Y., Ye, K., and Smith, E. P. (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics* **17**, 95–106.

- Evans, M. and Jang, G. H. (2011a). A limit result for the prior predictive applied to checking for prior-data conflict. *Statistics & Probability Letters* **81**, 1034–1038.
- Evans, M. and Jang, G. H. (2011b). Weak informativity and the information in one prior relative to another. *Statistical Science* **26**, 423–439.
- Evans, M. and Moshonov, H. (2006). Checking for prior-data conflict. *Bayesian Analysis* **1**, 893–914.
- Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics* **13**, 342–368.
- Gelman, A., Jakulin, A., Grazia, M. P., and Yu-Sung, S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* **2**, 1360–1383.
- Greenland, S. (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology* **35**, 765–775.
- Greenland, S. (2007). Bayesian perspectives for epidemiological research. II. Regression analysis. *International Journal of Epidemiology* **36**, 195–202.
- Greenland, S. and Mansournia, M. A. (2015). Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in Medicine* **34**, 3133–3143.
- Held, L., Sabanés Bové, D., and Gravestock, I. (2015). Approximate Bayesian model selection with the deviance statistic. *Statistical Science* **30**, 242–257.
- Held, U., Sabanés Bové, D., Steurer, J., and Held, L. (2012). Validating and updating a risk model for pneumonia - a case study. *BMC Medical Research Methodology* **12**, 99.
- Hoerl, A. E., Kennard, R. W., and Baldwin, K. F. (1975). Ridge regression: Some simulations. *Communications in Statistics: Theory & Methods* **4**, 105–124.

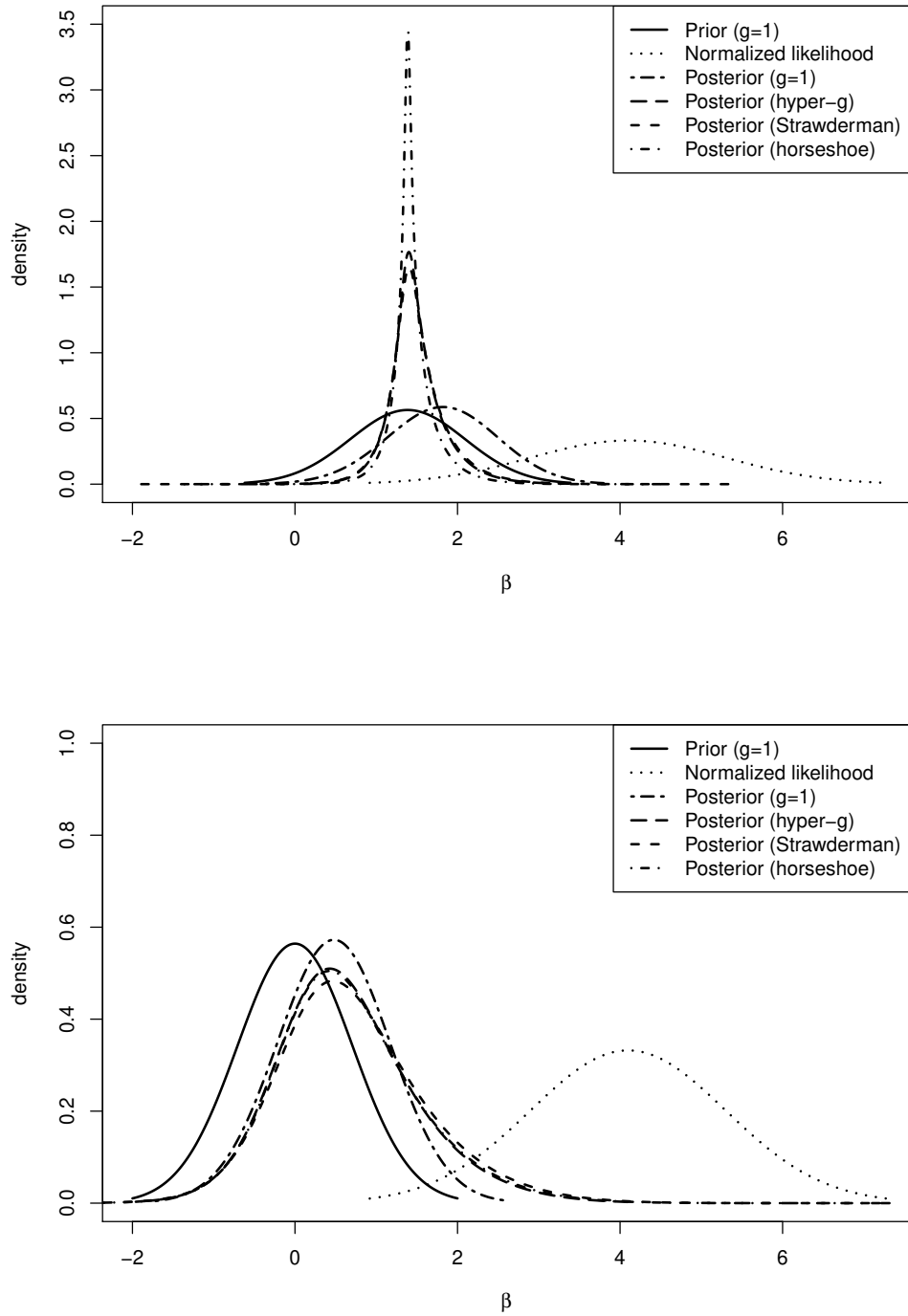
- Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science* **15**, 46–60.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of  $g$  priors for Bayesian variable selection. *Journal of the American Statistical Association* **103**, 410–423.
- Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)* **34**, 1–41.
- Marin, J.-M. and Robert, C. P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer texts in Statistics. Springer, New York.
- Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics and Data Analysis* **67**, 68–83.
- Miettinen, O. S., Flegel, K. M., and Steurer, J. (2008). Clinical diagnosis of pneumonia, typical of experts. *Journal of Evaluation in Clinical Practice* **14**, 343–350.
- Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). A note on the power prior. *Statistics in Medicine* **28**, 3562–3566.
- Neutra, R. R., Fienberg, S. E., Greenland, S., and Friedman, E. A. (1978). Effect of fetal monitoring on neonatal death rates. *New England Journal of Medicine* **299**, 324–326.
- O’Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements; Eliciting Experts’ Probabilities*. Wiley, Chichester.
- Roos, M. and Held, L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis* **6**, 259–278.
- Roos, M., Martins, T. G., Held, L., and Rue, H. (2015). Sensitivity analysis for Bayesian

- hierarchical models. *Bayesian Analysis* **10**, 321–349.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society - Series B* **71**, 319–392.
- Sabanés Bové, D. and Held, L. (2011). Hyper-g priors for generalized linear models. *Bayesian Analysis* **6**, 387–410.
- Schmidli, H., Gsteiger, S., Roychoudhury, S., O’Hagan, A., Spiegelhalter, D., and Neuen-schwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* **70**, 1023–1032.
- Spiegelhalter, D., Best, N., Carlin, B., and Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* **64**, 583–639.
- Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, New York.
- Sullivan, S. G. and Greenland, S. (2013). Bayesian regression in SAS software. *International Journal of Epidemiology* **42**, 308–317.
- West, M. (1987). On scale mixtures of normal distributions. *Biometrika* **74**, 646–648.
- Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Goel, P. K. and Zellner, A., editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, volume 6 of *Studies in Bayesian Econometrics and Statistics*, chapter 5, pages 233–243. North-Holland, Amsterdam.

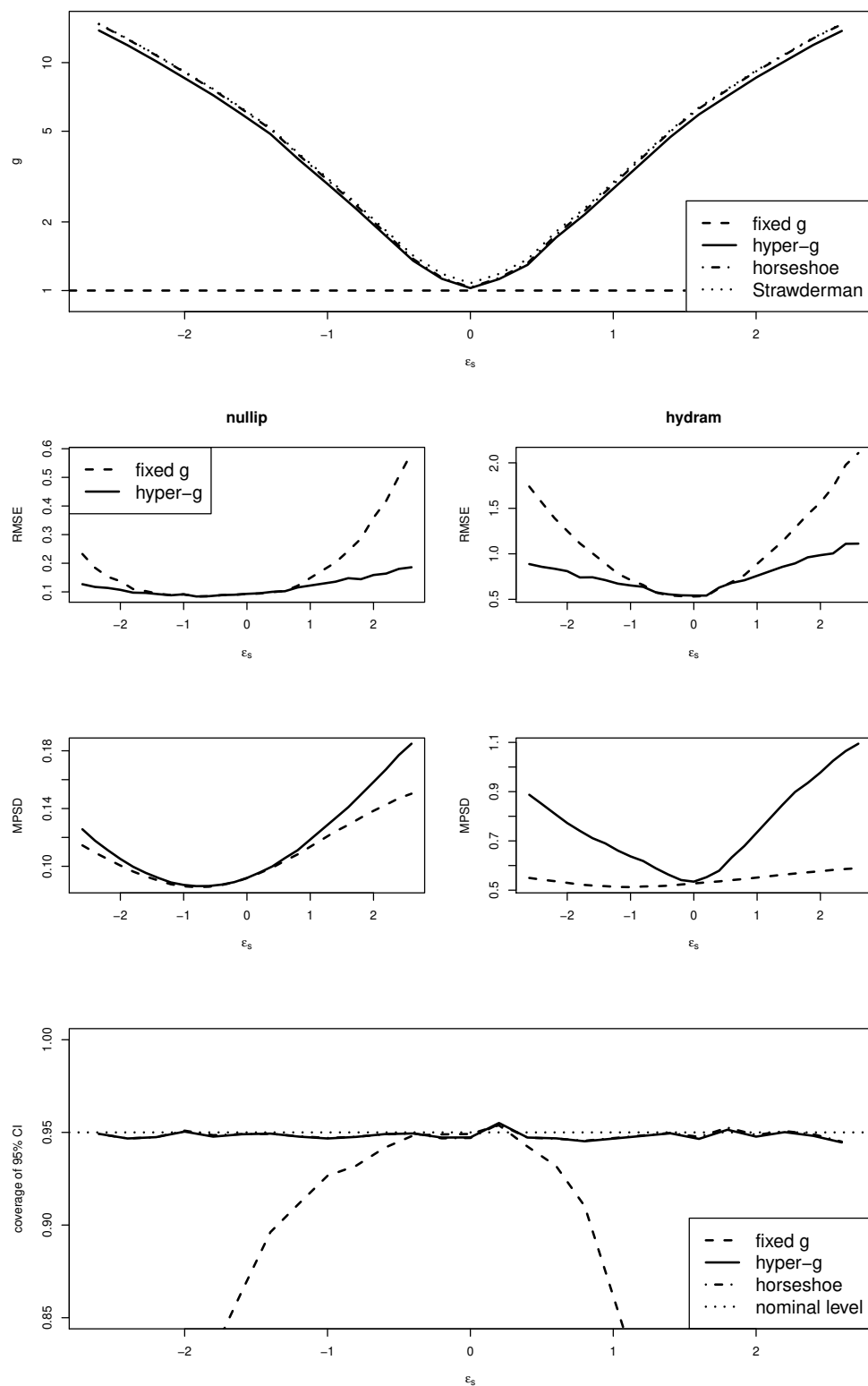


**Figure 1:** Hyper- $g$  prior and posterior density of  $g$  (left) and the corresponding  $t = g/(1 + g)$  (right) in the logistic regression example. Posterior 1 based on  $N(\nu_{SG}, g \Sigma)$  prior for  $\beta$ . Posterior 2 based on  $N(\mathbf{0}, g \Sigma)$  prior for  $\beta$ .

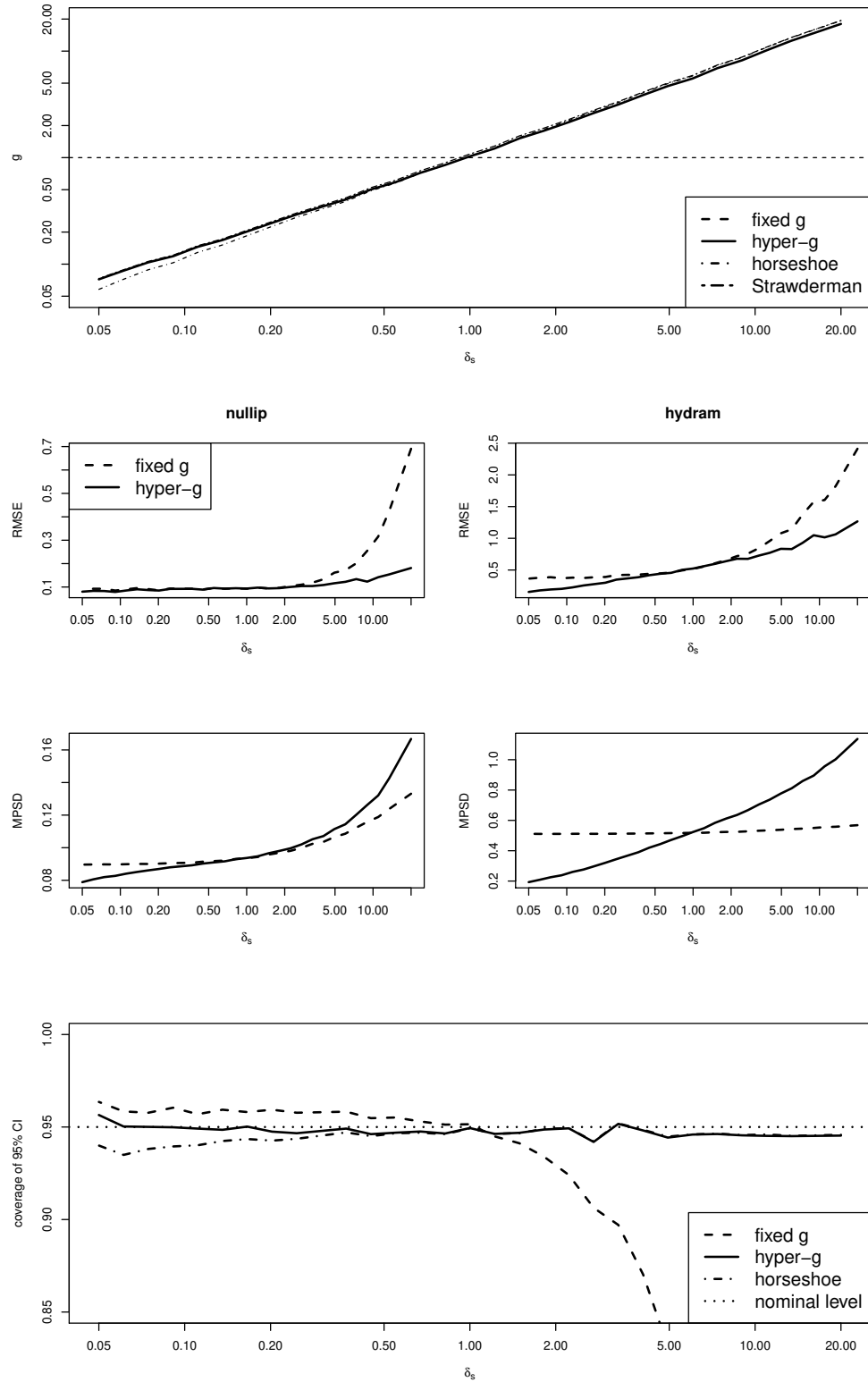




**Figure 2:** Posterior density of regression coefficient  $\beta_7$  for the variable hydram in the logistic regression example with fixed ( $g = 1$ ) and adaptive (hyper- $g$ , Strawderman and horseshoe) prior weighting. Top:  $N(\nu_{SG}, g \Sigma)$  prior. Bottom:  $N(\mathbf{0}, g \Sigma)$  prior.



**Figure 3:** Simulation I: Mean posterior median estimates of  $g$  (top row), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for two explanatory variables (nullip and hydram) (middle rows) and average coverage of 95% credible intervals across all explanatory variables (bottom row).



**Figure 4:** Simulation II: Mean posterior median estimates of  $g$  (top row), root mean squared error (RMSE) and mean posterior standard deviation (MPSD) for two explanatory variables (nullip and hydram) (middle rows) and average coverage of 95% credible intervals across all explanatory variables (bottom row).

	$\nu = \nu_{SG}$			$\nu = \mathbf{0}$		
	OR	95% CI	DIC	OR	95% CI	DIC
Likelihood	60	2.8 to 478.2		60	2.8 to 478.2	
g=1	6.1	1.6 to 22.8	183.4	1.6	0.4 to 6.3	188.3
Strawderman	4.3	2.3 to 11.1	180.5	1.9	0.4 to 16.1	188.8
Hyper-g	4.3	2.3 to 10.5	180.3	1.8	0.4 to 13.4	188.8
Horseshoe	4.1	2.6 to 8.4	179.6	1.8	0.4 to 14.4	188.9
block Hyper-g				5.5	0.4 to 95.5	182.6
Prior (g=1)	4.0	1.0 to 16.0		1.0	0.25 to 4.00	

Table 1: Odds ratio (OR) estimate and 95% credible interval for hydramnios coefficient with prior mean  $\nu_{SG}$  (left) or prior mean  $\mathbf{0}$  (right) together with DIC values for different priors on  $g$ . Shown are also the corresponding numbers based on the likelihood only (first row) and the prior only (last row).